Metadata Requirements Beyond CF

Karl Taylor

Program for Climate Model Diagnosis and Intercomparison
Lawrence Livermore National Laboratory

GO-ESSP Meeting

Didcot, England

6-8 June 2005

Outline

- Additional (higher level) metadata documenting model configuration and experiments performed
- IPCC requirements beyond CF that make files more “self-descriptive”

Information contained in IPCC model documentation:

- Enables a more comprehensive analysis and scrutiny of model behavior by a wider community of scientists
- Should speed advances in the understanding of model behavior

Organization of information in model documentation

I. **Model identity**

II. **How comprehensive is model?**

III. **List the community based projects that your modeling group has participated in (e.g., AMIP, C4MIP, PMIP, PILPS, etc.).**

IV. **Component model characteristics (of current IPCC model version):**

   A. Atmosphere
   B. Ocean
   C. Sea ice
   D. Land surface and ice sheets
   E. Coupling details

V. **Simulation Details (report separately for each IPCC simulation)**

SAMPLE MODEL DOCUMENTATION

Model Information of Potential Use to the IPCC Lead Authors and the AR4.

MRI-CGCM2.3.2

Meteorological Research Institute, Japan

18 February 2005

I. Model identity:

A. Institution, sponsoring agency, country
   Meteorological Research Institute, Japan Meteorological Agency, Japan

B. Model name (and names of component atmospheric, ocean, sea ice, etc. models)
   MRI-CGCM2.3.2 (TAR = MRI-CGCM2.0)

C. Vintage (i.e., year that model version was first used in a published application)
   2003

D. General published references and web pages
   Yukimoto et al. (2001), *Pap. Meteor. and Geophys.*, 51, 47-88.
   http://www.mri-jma.go.jp/Dep/cl/cl4/publications/yukimoto_pap2001.pdf
   The above document is for the TAR version (MRI-CGCM2.0).

E. References that document changes over the last ~5 years (i.e., since the IPCC TAR) in the coupled model or its components. We are specifically looking for references that document changes in some aspect(s) of model performance.
   The major changes for the current version are documented in Yukimoto and Noda (2002, *CGER's supercomputer activity report*, 10, 37-44, NIES, Japan, http://www.mri-jma.go.jp/Dep/cl/cl4/publications/yukimoto_CGER2002.pdf) for the former version (MRI-CGCM2.2). The changes for the current version (MRI-CGCM2.3.2) from the former version, which are small, have not published yet. It is in preparation for submission to a journal.

F. IPCC model version's global climate sensitivity ($KW^{-1}m^2$) to increase in CO_2 and how it was determined (slab ocean expt., transient expt--Gregory method, ±2K Cess expt., etc.)
   The climate sensitivity is 0.86 $K/(Wm^{-2})$, which is estimated from a slab ocean 2xCO2 equilibrium experiment.

G. Contacts (name and email addresses), as appropriate, for:

1. coupled model : Seiji Yukimoto, yukimoto@mri-jma.go.jp
2. atmosphere : Takao Uchiyama, tuchiyam@mri-jma.go.jp
3. ocean : Seiji Yukimoto, yukimoto@mri-jma.go.jp
4. sea ice : Seiji Yukimoto, yukimoto@mri-jma.go.jp
5. land surface : Masahiro Hosaka, mhosaka@mri-jma.go.jp
6. vegetation : Masahiro Hosaka, mhosaka@mri-jma.go.jp
7. other?

II. Besides atmosphere, ocean, sea ice, and prescription of land/vegetated surface, what can be included (interactively) and was it active in the model version that produced output stored in the PCMDI database?

A. atmospheric chemistry?
   Not included
B. interactive biogeochemistry?
   Not included
C. what aerosols and are indirect effects modeled?
   Only direct effect of sulfate aerosol
D. dynamic vegetation?
   Not included
E. ice-sheets?
   Not included

III. List the community based projects (e.g., AMIP, C4MIP, PMIP, PILPS, etc.) that your modeling group has participated in and indicate if your model results from each project should carry over to the current (IPCC) version of your model in the PCMDI database.

AMIP: different version (MRI/JMA98, Shibata et al. 1999, *Pap. Meteor. Geophys*, 50, 15-53)
PMIP: the same as for AMIP
SMIP: the same as for AMIP

IV. Component model characteristics (of current IPCC model version):

A. Atmosphere

General published documentation of the atmospheric component:

Shibata et al. (1999, *Pap. Meteor. Geophys*, 50, 15-53)
http://www.mri-jma.go.jp/Dep/cl/cl4/publications/shibata_pap1999.pdf

1. resolution
   - horizontal resolution: T42 (approx. 2.8 degrees)
2. numerical scheme/grid (advective and time-stepping schemes; model top; vertical coordinate and number of layers above 200 hPa and below 850 hPa)
   - numerical scheme: spectral transform method
   - time-stepping scheme: leap-frog, semi-implicit with Asselin time filter
   - model top: 0.4 hPa
   - vertical coordinate: sigma-pressure hybrid, 30 layers (16 layers above 200 hPa, 5 layers below 850 hPa)
3. list of prognostic variables (be sure to include, as appropriate, liquid water, chemical species, ice, etc.). Model output variable names are not needed, just a generic descriptive name (e.g., temperature, northward and eastward wind components, etc.)
   - velocity (velocity potential and streamfunction)
   - temperature
   - specific humidity
4. name, terse descriptions, and references (journal articles, web pages) for all major parameterizations. Include, as appropriate, descriptions of:
   a. clouds
      diagnostic clouds based on function of relative humidity
      Yukimoto et al. (2001, *Pap Meteor and Geophys*, 51, 47-88)
      http://www.mri-jma.go.jp/Dep/cl/cl4/publications/yukimoto_pap2001.pdf
   b. convection
      Prognostic Arakawa-Schubert
      Randall and Pan (1993)
   c. boundary layer
      turbulent closure level 2
      Mellor and Yamada (1974, *J. Atmos. Sci.*, 31, 1791-1806)

d. SW, LW radiation

SW: delta-two-stream

Shibata and Uchiyama (1992, *J. Meteor. Soc. Japan*, 70, 1097-1109)

LW: Multi-parameter random model

Shibata and Aoki (1989, *J. Geophys. Res.*, 94, 14923-14943)

e. any special handling of wind and temperature at top of model

Rayleigh friction with hyperbolic tangent profile

Shibata et al. (1999, *Pap. Meteor. Geophys*, 50, 15-53)

B. Ocean

1. resolution

   2.5 degree (longitude) x 2.0 degree (latitude, poleward of 12S and 12N) ~ 0.5 (4S-4N)

2. numerical scheme/grid, including advection scheme, time-stepping scheme, vertical coordinate, free surface or rigid lid, virtual salt flux or freshwater flux

   - numerical scheme/grid: Arakawa B-grid
   - time-stepping scheme: leap-frog and Matsuno scheme
   - vertical coordinate: z-coordinate
   - surface condition: regid lid
   - freshwater flux: virtual salt flux

3. list of prognostic variables and tracers

   - velocities (eastward and northward)
   - temperature
   - salinity

4. name, terse descriptions, and references (journal articles, web pages) for all parameterizations. Include, as appropriate, descriptions of:

   a. eddy parameterization

   Laplacian-type horizontal diffusion + GM paramererization (Gent and McWilliams, 1990, *J. Phys. Oceanogr.*, 20, 150-155)

   b. bottom boundary layer treatment and/or sill overflow treatment

   Not included.

   c. mixed-layer treatment

   turbulent closure level 2

   Mellor and Yamada (1974, *J. Atmos. Sci.*, 31, 1791-1806)

Mellor and Durbin (1975, *J. Phys. Oceanogr.*, 5, 718-728)

d. sunlight penetration
   Sunlight penetrates with a constant decaying rate of 10 meters e-folding depth.
e. tidal mixing
   Not included.
f. river mouth mixing
   Larger vertical diffusivities in the upper 30m are applied at the Amazon's river mouth.
g. mixing isolated seas with the ocean
   Not included. (There is no isolated seas.)
h. treatment of North Pole "singularity" (filtering, pole rotation, artificial island?)
   Fourier filtering at northward of 82N

C. sea ice

1. horizontal resolution, number of layers, number of thickness categories
   - horizontal resolution: 2.5 degrees (longitude) x 2.0 degrees (latitude)
   - number of layers: 1 layer
   - no thickness categories
2. numerical scheme/grid, including advection scheme, time-stepping scheme,
   - numerical scheme/grid: Arakawa B-grid
   - time-stepping scheme: forward differencing
3. list of prognostic variables
   - thickness
   - compactness
4. completeness (dynamics? rheology? leads? snow treatment on sea ice)
   Thickness and compactness (leads) are advected by a function of ocean surface current. Diffusion is also applied for thickness and compactness. Rheology is not included. Thermo-dynamical processes of snow on sea ice are treated.
5. treatment of salinity in ice
   Salinity of sea ice is fixed at a constant value (4 psu).
6. brine rejection treatment
   No brine rejection from sea ice is treated, since the salinity of sea ice is constant. Salinity flux into ocean is accounted only for frazil ice formation and melt/freeze of sea ice.
7. treatment of the North Pole "singularity" (filtering, pole rotation, artificial island?)
   There is no special treatment of the North Pole for sea ice, though Fourier filtering is applied for ocean.

D. **land surface and ice sheets** (some of the following may be omitted if information is clearly included in cited references.

1. resolution (tiling?), number of layers for heat and water
   - horizontal resolution: T42 Gaussian grid (~300km, same as the atmosphere)
   - number of layers: 3 layers (for both heat and water)
2. treatment of frozen soil and permafrost
   Frozen soil water is treated as a prognostic variable. Where the soil layer is year-around frozen can be defined as permafrost.
3. treatment of surface runoff and river routing scheme
   River routing scheme is included. Surface runoff occurs when the precipitation rate is larger than the maximum (saturated) infiltration ratio. Surface runoff is discharged into ocean at the river mouth via the river routing.
4. treatment of snow cover on land
   One snow layer. The energy and water mass are conserved in the thermo-dynamical processes of snow.
5. description of water storage model and drainage
   Prognostic surface temperature and water storage for four major lakes are modeled, but drainage from the lakes is not treated. The drainage from the bottom soil layer is included, and the amount is evaluated from the vertical gradient of hydraulic potential. The drainage water is also discharged through the river route.
6. surface albedo scheme
   The soil skin albedo depends on the vegetation type. The snow skin albedo depends on the snow temperature. In case that the amount of snow is small, the partial snow is considered in the estimation of the skin albedo. The radiation transfer between canopy top and skin is calculated, then albedo for the surface or canopy top is evaluated.
7. vegetation treatment (canopy?)
   Canopy and grass. There are 13 vegetation types, and the vegetation parameters depend on the type and month of the year.
8. list of prognostic variables
   Temperature of canopy, skin(including grass), 3 soil layers. The moisture of canopy, grass, 3 soil layers.
9. ice sheet characteristics (How are snow cover, ice melting, ice accumulation, ice dynamics handled? How are the heat and water fluxes handled when the ice sheet is melting?)
   The ice sheet is one of the vegetation types. It is treated as white soil. On the “soil”, the snow accumulates and melts. When the mass of the snow is more than some threshold (10 m), the mass is passed to the river routing model, and it is immediately discharged as iceberg to the ocean. In any case, the handling of the heat and water fluxes is the same as those over snow layer of other vegetation type.

E. coupling details

1. frequency of coupling

24 hours

2. Are heat and water conserved by coupling scheme?

   Yes.

3. list of variables passed between components:
   a. atmosphere – ocean
      - heat flux (sensible heat + latent heat + long wave radiation)
      - solar radiative flux
      - freshwater mass flux
      - wind stress (eastward and northward)
   b. atmosphere – land
      - sensible heat flux
      - latent heat flux
      - long wave radiation
      - solar radiation
      - precipitation
      - evaporation/sublimation
      - wind stress (eastward and northward)
   c. land – ocean
      - freshwater mass flux (river discharge)
      - ice mass flux (iceberg discharge)
   d. sea ice – ocean
      - heat flux
      - salt flux
      - frazil ice mass
      - ocean surface velocities (eastward and northward)
   e. sea ice – atmosphere
      - sensible heat flux
      - latent heat flux
      - long wave radiation
      - solar radiation
      - precipitation
      - evaporation/sublimation
      - wind stress (eastward and northward)

4. Flux adjustment? (heat?, water?, momentum?, annual?, monthly?).
   Monthly climatological flux adjustment for heat, water and momentum (only 12S-12N) is used.

V. Simulation Details (report separately for each IPCC simulation contributed to database at PCMDI):

A. PIcntrl

The “PIcntrl” is pre-industrial simulation used for the reference to the historical (20C3M) and SRES experiments, and also provides initial states for the “20C3M”.

The initial state of “PIcntrl” is taken from the final state of a 450-year pre-industrial spin-up (PIspup) which initiated from the end of present-day spin-up run (428 year length, see comments on PDcntrl). The PIspup is fully coupled mode time integration, and is long enough so there is little climatic drift in PIcntrl at least in the upper ocean.

The forcing agents of the “PIspup” and “PIcntrl” experiments include greenhouse gases, sulfate aerosol (only natural source) direct effects and solar forcing, and is fixed at levels of year 1850 (CO_2 = 290 ppmv, CH_4 = 792 ppbv, N_2O = 285 ppbv, solar constant = 1366.0 Wm^{-2}, refer http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).

B. PDcntrl

The “PDcntrl” is a present-day control simulation, and is used for reference to the 1%/year CO_2 increase experiments (1%to2x and 1%to4x).

The initial state of the “PDcntrl” is taken from the end of the present-day spin-up (PDspup) of 428-year length, which is initiated from Levitus climatological ocean state. The “PDspup” is executed with fully coupled mode, and consists of a 334-year restore stage and a 95 year flux-correct stage. In the former stage, the temperature and salinity is restored to the observational climatology. In the latter stage, the ocean is driven by atmospheric fluxes and flux correction that is made from the climatology of restoring terms in the last part of the restore stage.

The forcing agents of the “PDspup” and “PDcntrl” experiments include greenhouse gases, sulfate aerosol direct effects and solar forcing, and is fixed at present-day (CO_2 = 348 ppmv, CH_4 = 1650 ppbv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}, refer http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).

C. 20C3M

This experiment consists of 5-member ensemble simulations of the 20th Century climate (starting from mid-19th Century).

The initial state of each member simulation is taken from the different states of the pre-industrial control (PIcntrl) experiment at 1, 51, 101, 151, and 200 years. The forcing agents of the experiment include greenhouse gases (CO_2, CH_4, N_2O and CFCs), sulfate aerosol direct effects, volcanoes and solar forcing. The distributions and temporal changes are provided in http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html .

D. SRESA1B

This experiment consists of 5-member ensemble simulation of future climate projection for IPCC SRES-A1B scenario.
Each member simulation starts from the state of respective year 1990 in the “20C3M” ensemble, although the “20C3M” are extended until year 2000.
The forcing agents of the experiment includes greenhouse gases CO_2, CH_4, N_2O and CFCs) and sulfate aerosol direct effects, which are based on IPCC SRES-A1B for 1990-2100. The distributions and temporal changes are provided in http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html
One of the members (run-1) is extended until 2300 with holding the concentrations fixed at 2100 level, and the rest of the runs end at 2100.

E. SRESA2

This experiment consists of 5-member ensemble simulation of future climate projection for IPCC SRES-A2 scenario.
Each member simulation starts from the state of respective year 1990 in the “20C3M” ensemble, although the “20C3M” are extended until year 2000.
The forcing agents of the experiment includes greenhouse gases CO_2, CH_4, N_2O and CFCs) and sulfate aerosol direct effects, which are based on IPCC SRES-A2 for 1990-2100. The distributions and temporal changes are provided http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html

F. SRESB1

This experiment consists of 5-member ensemble simulation of future climate projection for IPCC SRES-B1 scenario.
Each member simulation starts from the state of respective year 1990 in the “20C3M” ensemble, although the “20C3M” are extended until year 2000.
The forcing agents of the experiment include greenhouse gases (CO_2, CH_4, N_2O and CFCs) and sulfate aerosol direct effects, which are based on IPCC SRES-B1 for 1990-2100. The distributions and temporal changes are provided in http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html
One of the members (run-1) is extended until 2300 with holding the concentrations fixed at 2100 level, and the rest of the runs end at 2100.

G. 1%to2x

This experiment simulates transient climate response to 1%/year increase of CO_2 concentration.

The initial state is identical to the beginning of "PDcntrl" simulation, which is taken from the end of the 428-year present-day spin-up (PDspup).
The CO_2 concentration increases (starting from 348 ppmv) with 1%/year compound rate and reaches doubled (696 ppmv) at 70th year, and is fixed doubled after that. The other forcing agents are same as in the "PDcntrl" simulation (e.g., CH_4 = 1650 ppmv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}).

H. 1%to4x

This experiment simulates transient climate response to 1%/year increase of CO_2 concentration.
The initial state is identical to the beginning of "PDcntrl" simulation, which is taken from the end of the 428-year present-day spin-up (PDspup).
The CO_2 concentration increases (starting from 348 ppmv) with 1%/year compound rate and reaches doubled (1392 ppmv) at 140th year, and is fixed quadrupled after that. The other forcing agents are same as in the "PDcntrl" simulation (e.g., CH_4 = 1650 ppmv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}).

I. Slabcntl

This simulation is a control simulation with the slab ocean model, which is used for reference to the equilibrium response experiment (2xCO2).
Starting from the initial state with observed SST climatology (Levitus, 1994), the model is spun-up with coupled mode by restoring the slab temperature to the observed SST climatology to make 'Q-flux'. By using the 'Q-flux' the model is run long enough (250 years) to reach a complete equilibrium and to produce stable statistics for the sampling period of the last 50 years.
The forcing agents are same as in the "PDcntrl" simulation (CO_2 = 348 ppmv, CH_4 = 1650 ppbv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}, http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).

J. 2xCO2

This is a simulation of equilibrium response to an instantaneous doubling of CO_2 (696 ppmv) with the slab ocean model.
The initial state is identical to the "Slabcntl" simulation. It's run is long enough (250 years) to reach a complete equilibrium and to produce stable statistics for the last 50 years for the sampling period.
The other forcing is same as in the "Slabcntl" experiment (CH_4 = 1650 ppmv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}, http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).

K. Commit

This simulates committed climate change by the forcing in the past, assuming all the anthropogenic forcing is fixed at year 2000 levels (CO_2 = 393.7 ppmv, CH_4 = 1738 ppbv, N_2O = 315.5 ppbv, solar constant = 1367.0 Wm^{-2}, http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).
The initial state is taken from the end of the run-1 of “20C3M”.

L. AMIP

This is an ‘AMIP type’ simulation with the atmospheric component of the MRI-CGCM2.3.2 which is identical to other IPCC experiments. The atmospheric initial state, anthropogenic forcing and solar forcing are same as in the “PDcntrl” experiment (CO_2 = 348 ppmv, CH_4 = 1650 ppbv, N_2O = 306 ppbv, solar constant = 1367.0 Wm^{-2}, refer http://www.mri-jma.go.jp/Dep/cl/cl4/IPCC-AR4/simulations.html for detail).
The simulation period is through 1978 to 2002, for which the ‘AMIP II SST and sea ice boundary condition data set’ is used.

Beyond CF requirements: The CF-conventions specify the "syntax" of self-describing files, but the IPCC requirements are stricter.

- Each file must contain only a single output field.
- For data that are a function of longitude and latitude, only grids representable as a Cartesian product of longitude and latitude axes are allowed.
- Most atmospheric fields that are functions of the vertical coordinate must be interpolated to standard pressure levels
- The names of output netCDF files must begin with the 'output variable name' listed in the IPCC Standard Output list of variables, followed by an underscore and then the table number (e.g., tas_O1, psl_A1, tas_A4, etc.).

http://www-pcmdi.llnl.gov/ipcc/IPCC_output_requirements.htm

Beyond CF requirements (cont.)

- The units required for the output fields are given in the IPCC Standard Output tables.
- The positive direction of vertical fluxes must be consistent with that specified in the IPCC Standard Output table "CF standard name".
- The order of array dimensions must be: time, level, latitude, longitude.
- Data must be stored south to north and west to east , starting with the first grid point ≥ 0° E
- If there is a vertical coordinate, data must be stored starting with the grid point nearest the surface
- Coordinates bounds are usually required.

http://www-pcmdi.llnl.gov/ipcc/IPCC_output_requirements.htm

Sample entries in IPCC standard output specifications:

Table A1a: **Monthly-mean 2-d atmosphere or land surface data (longitude, latitude, time:month).**

	CF standard_name	**output variable name**	**units**	**notes**
1	air_pressure_at_sea_level	psl	Pa	
8	surface_snow_thickness	snd	m	this thickness when multiplied by the average area of the grid cell covered by snow yields the time-mean snow volume. Thus, for time means, compute as the weighted sum of thickness (averaged over the snow-covered portion of the grid cell) divided by the sum of the weights, with the weights equal to the area covered by snow. report as 0.0 in snow-free regions.
15	surface_temperature	ts	K	"skin" temperature (i.e., SST for open ocean)
16	surface_air_pressure	ps	Pa	*not* mean sea-level pressure
19	atmosphere_water_vapor_content	prw	kg m^{-2}	vertically integrated through the atmospheric column
21	surface_runoff_flux	mrros	kg m^{-2} s^{-1}	compute as the total surface runoff leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.
22	runoff_flux	mrro	kg m^{-2} s^{-1}	compute as the total runoff (including "drainage" through the base of the soil model) leaving the land portion of the grid cell divided by the land area in the grid cell; report as "missing" or 0.0 where the land fraction is 0.

http://www-pcmdi.llnl.gov/ipcc/standard_output.html

Climate Model Output Rewriter (CMOR)

- A FORTRAN 90 library for rewriting model output and ensure compliance with the IPCC requirements.
- Relies on an input file to supply much of the “standard metadata” associated with the IPCC standard output fields.
- The CMOR input files can be reconfigured to change the requirements and meet the needs of other model intercomparison projects.
- http://www-pcmdi.llnl.gov/software/cmor/cmor_users_guide.pdf

CMOR capabilities include data manipulation to make it consistent with specifications in the CMOR input tables

- Reordering of dimensions
- Reversing order of coordinate values
- Automatic units conversion and appropriate data scaling (relying on udunits and cdms)
- Automatic type conversion of data (e.g., “single” to “double precision”)
- Automatic substitution of standard “missing_value” for user’s missing_value.

CMOR checks input for consistency with requirements

- Are required pressure levels included?
- Are data values within a reasonable range? (traps sign and units errors).
- Are standard identifiers (e.g., variable names, experiment i.d.'s) in the list of acceptable values?
- Are coordinate values monotonic?
- Are the units consistent with the variable definition?

CMOR input tables provide many of the required attributes including:

- Global attributes: title, conventions, history
- Variable attributes: standard_name, units, long_name, cell_methods, coordinates, history
- Axis attributes: axis, coordinates, bounds, positive, formula_terms

CMOR limitations:

- All longitude x latitude grids must be represented as a Cartesian product of longitude and latitude axes
- Each file contains only a single output variable
- Cannot interpolate data to a new grid or in the vertical
- Except for all variants of hybrid-sigma coordinates, the vertical coordinate must be independent of horizontal position.

Summary

- Considerable model documentation is available, but not in the most convenient form.
- IPCC metadata requirements facilitate automated analysis of model output from a large collection of models and experiments.
- CMOR facilitates compliance with IPCC requirements (but is not mandated)